Road network selection using an extended stroke-mesh combination algorithm

Benz, Stefan A ; Weibel, Robert

DOI: https://doi.org/10.1080/15230406.2014.928482

Originally published at:
Benz, Stefan A; Weibel, Robert (2013). Road network selection using an extended stroke-mesh combination algorithm. In: 16th ICA Workshop on Generalisation and Multiple Representation, Dresden, Germany, 23 August 2013 - 24 August 2013, Commission on Generalisation and Multiple Representation.
DOI: https://doi.org/10.1080/15230406.2014.928482

Road Network Selection Using an Extended Stroke-Mesh Combination Algorithm

S.A. Benz[1], R. Weibel[1]

[1]Department of Geography, University of Zurich, Winterthurerstr. 190, 8057 Zürich, Switzerland
Email: {stefan.benz | robert.weibel}@geo.uzh.ch

1. Introduction

The road network is an essential part of topographic maps and databases and its generalisation is considered to be complex. Foremost, roads are connected to each other and form a coherent network. Deletion of roads without consideration of the entire network results in loss of connectivity (Chaudhry and Mackaness 2005).

The selection of road segments usually precedes other generalisation operators. It aims to reduce the level of detail in the road network by choosing the relevant road segments but also maintaining the main characteristics and structure (Touya 2010). There exist several algorithms that perform the selection automatically. Among those, the stroke-based approach (Thomson and Richardson 1999) and the mesh-based approach (Chen et al. 2009) have proven to produce promising results. Recently, an algorithm has been introduced that combines both the stroke- and the mesh-based approach in an integrated concept (Li and Zhou 2012). However, this algorithm has not been tested for many test cases yet, especially not for heterogeneous regions.

This paper reports on a research project that is pursued in collaboration with swisstopo, the national mapping agency of Switzerland. Currently, the selection at swisstopo is done manually but they have shown interest in using an automated approach. Therefore, an analysis has been conducted of how well the three approaches (stroke-based, mesh-based and the combination thereof) perform for a target scale of 1:50’000, with respect to requirements set forth by swisstopo experts that the result must fulfil. The aim set by swisstopo is to select 70 % of the features from the source database for the target scale. The 1:50’000 topographical maps from swisstopo are derived from the TLM3D (Topographic Landscape Model 3D), which thus serves as a basis for this project. It is an extremely detailed, dense and large-scale spatial database (corresponding to the scale range of 1:5’000 – 1:25’000) that covers Switzerland with a high spatial resolution. Four test areas with different conditions (urban, rural, mixed, alpine) are being used to evaluate the algorithms.

After the implementation of the three algorithms, a first round of experiments showed that the integrated stroke-mesh algorithm produces the best results. However, several difficulties remained and not all of the requirements could be fulfilled. Therefore, an analysis was carried out as to how additional constraints and extensions could improve the results.

The contributions of this work are the following: (1) a thorough empirical analysis of three algorithms for road network selection (stroke-based, mesh-based, and integrated stroke-mesh algorithms); (2) five extensions that address critical points of the integrated stroke-mesh algorithm and lead to a significant improvement of the results.

2. State of the art

Liu et al. (2010) divide automatic road network selection into three groups: (1) semantic-based selection, (2) graph-based selection and (3) stroke-based selection.

The first group is based on semantic attributes (e.g. road class). Roads are ordered according to their relative importance of attributes and the selection is based on this order. These methods are insufficient due to the neglect of geometrical and topological constraints.

The graph-based methods treat road networks as graphs and use pattern detection algorithms (Yang et al. 2011) or concepts, such as shortest/best path or minimum-spanning-trees, which serve as the basis for the selection (Mackaness and Beard 1993, Mackaness 1995). A special group of algorithms uses the dual graph approach, where nodes represent roads and edges represent intersections of roads, respectively. Dual graphs are often used to compute centrality measures, such as degree, closeness or betweenness as a measure of importance for roads (Jiang and Claramunt 2004).

Stroke-based selection is based on the principle of "good continuation". A stroke is a chain of road segments with continuous curvature (Thomson and Richardson 1999). Strokes are ordered according to some predefined rules (e.g. length) and the selection is conducted using strokes of higher order.

Agent-based methodologies provide a further approach for road network selection. For instance, Morisset and Ruas (1997) measured the importance of roads based on how much they are likely to be used by means of an agent-based simulation.

Another concept is the mesh-based approach (Chen et al. 2009). A mesh is a closed region that is bounded by several road segments (Figure 1). The selection is based on the identification of meshes with a high mesh density (ratio of the perimeter and the area of the mesh). In Figure 1, the mesh with the highest density (mesh number 3) is treated first. Its bounding segments are ordered according to their relative importance and the least important segment (red) is eliminated first. The remaining segments are merged with the adjacent mesh (mesh number 2), thus forming a new mesh with a lower mesh-density. This process is repeated until all meshes have a mesh-density smaller than a predefined threshold or until the desired amount of segments has been eliminated.

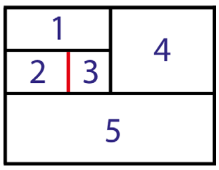

Figure 1. A road network with five meshes.

3. Integrated Stroke-Mesh Approach

Both the stroke-based and mesh-based approach cannot meet the requirements defined by swisstopo experts. Figure 2 provides a snippet of the result of each algorithm. The main problem with the stroke-based approach is the fact that the local connectivity cannot be maintained, i.e. roads are disconnected into separate parts and new dead-end roads appear (red circles in Figure 2b). The major flaw of the mesh-based approach (Figure 2c) is the fact that it does not handle linear segments, i.e. segments that are not a boundary of a mesh. The integrated approach introduced by Li and Zhou (2012) has the advantage that no new dead-end roads are generated and that segments not being a side of a mesh are handled, too (blue ellipses in Figure 2d). Therefore, it overcomes the main difficulties of the stroke- and the mesh-based approach.

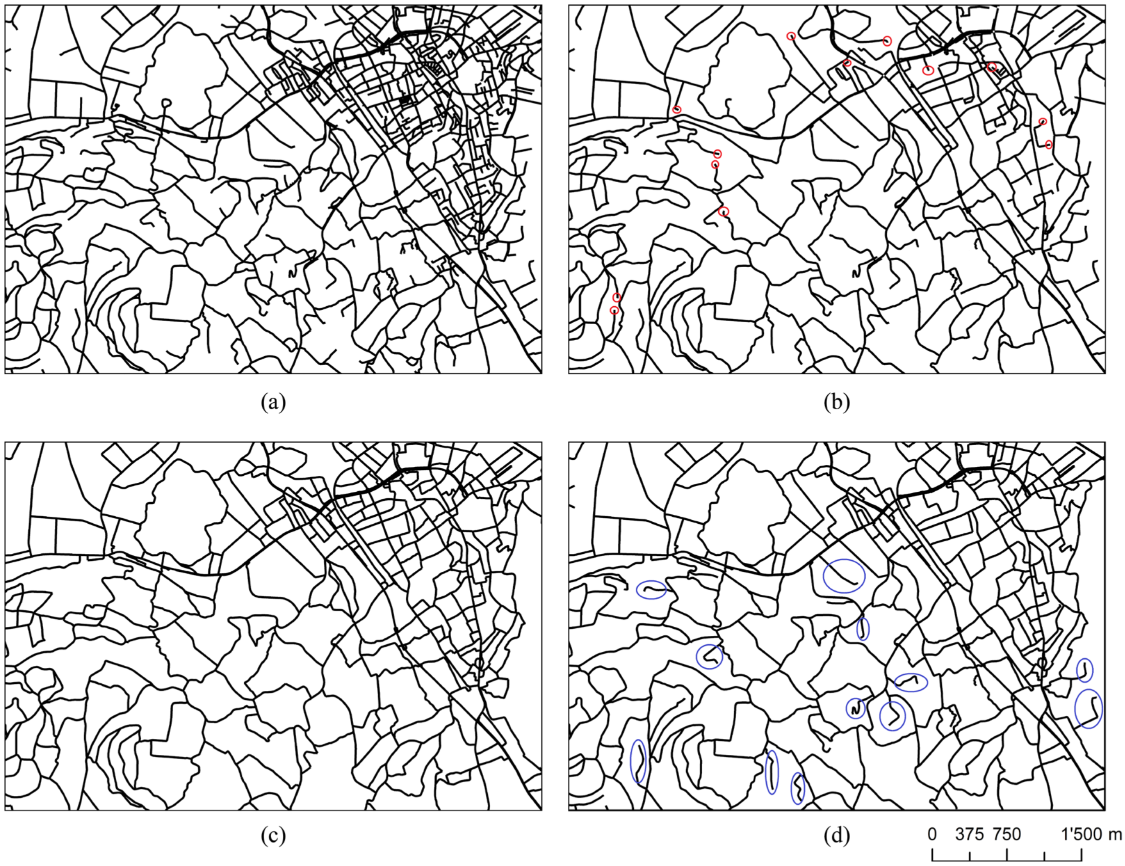

Figure 2. Road network snippet of the source database (a) and selection results using the stroke-based (b), the mesh-based (c) and the integrated approach (d). Data: TLM3D © swisstopo.

In order to understand the principle of the integrated approach, one has to differentiate between what Li and Zhou (2012) define as *areal* and *linear* segments, respectively. An areal segment is a segment that forms the boundary of a mesh, as is each segment in Figure 1. A linear segment is a segment that does not belong to any side of a mesh, such as a dead-end. Figure 3 provides an example of a road network with linear segments only.

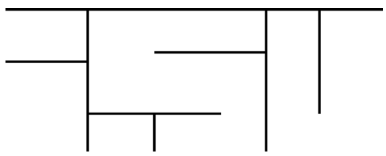

Figure 3. A road network with linear segments only.

Li and Zhou (2012) have shown that the stroke-based approach performs best for patterns with linear segments and the mesh-based approach for patterns with areal segments, respectively. Therefore, the integrated approach handles linear segments with the stroke-based approach and areal segments with the mesh-based approach.

A detailed description is provided in Li and Zhou (2012). The remainder of this section summarises the basic principle (two main steps) of the integrated stroke-mesh approach necessary to understand the extensions made in this project.

Basically, in the first step, all areal patterns (characterized by connected areal segments as shown in Figure 1) are thinned out using the mesh-based approach.

In the second step, the linear patterns (made up of linear segments only as shown in Figure 3) are handled using a tree hierarchy of strokes (Figure 4). First, road segments in a linear pattern are concatenated into strokes (strokes 1-5). These strokes are then placed into a tree data structure. The root node defines the starting point for the tree construction. Let us assume it is stroke 1. Stroke 3 directly connects to stroke 1, thus is a child-node of stroke 1. Stroke 2 and stroke 5 both connect to stroke 3 and thus are both children-nodes of stroke 3. Finally, stroke 4 connects to stroke 5 and thus is a child-node of stroke 5.

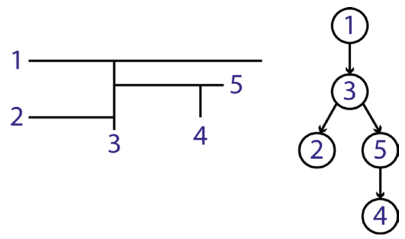

Figure 4. Linear pattern of strokes and the corresponding tree.

The tree hierarchy is used in the selection process of linear patterns. If such a hierarchy is connected to an areal segment that was *retained* in the first step of the integrated approach, the stroke that is connected to this areal segment builds the root node and all other strokes form children- and grandchildren-nodes. The traversal of the tree starts at the root. If the length of the currently visited stroke is longer than a specific threshold, the stroke is selected. The traversal then continues according to the same principle from this parent stroke to its children strokes, until the currently visited stroke has no more child stroke, or its length is shorter than a predefined threshold.

4. Problems of the integrated approach and proposed solutions

Although the integrated approach produces better results than the purely stroke- or mesh-based approach (Figure 2), several problems remain. After describing these problems and possible solutions thereof, an example of a result of the extended integrated approach (Figure 11) is provided.

4.1 Traversal of the tree data structure

Li and Zhou (2012) use a *top-down approach* in the traversal of the tree hierarchy for linear segments. However, this is problematic. Figure 5 shows a snippet of a road network with a linear pattern (a short stroke 1 and a set of longer strokes 2, 3 and 4) connected to an areal segment. Stroke 1 is the root node, whereas the other strokes form child and grandchild nodes. Using a top-down approach, one starts at stroke 1. If it is longer than the threshold, it is retained, otherwise it is eliminated and the traversal stops. The problem is that the other strokes are not considered if stroke 1 is shorter than the threshold, even though they could be significantly longer than the threshold. Therefore, we propose using a *bottom-up approach* where the traversal starts with the leaf nodes of the tree. If a stroke is longer than the threshold, the stroke itself *and* all its parent and grandparent strokes are retained, even if they are shorter than the threshold.

This ensures that all strokes longer than the threshold are selected, even if they appear at the bottom of the tree hierarchy, and it ensures that these strokes are connected to the network through their parent and grandparent strokes.

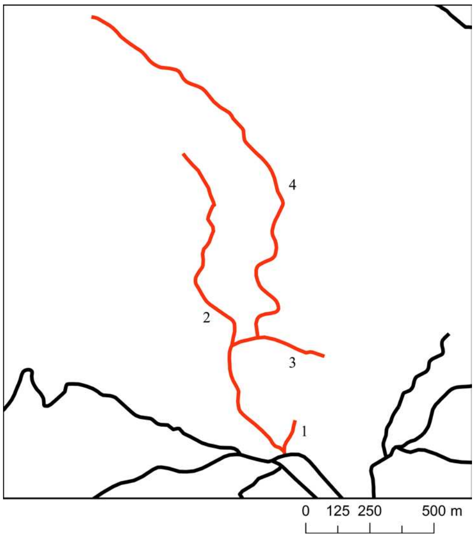

Figure 5. Problem caused by the top-down tree traversal. Data: TLM3D © swisstopo.

4.2 Unconnected linear patterns

Another problem is the fact that a linear pattern is only considered *if* it is connected to a mesh that was *retained* in the first step. Figure 6 (left) is an example where several linear patterns are connected to a dense mesh (green), and thus a mesh that is likely *not* to be retained. The consequence is that the large linear pattern within the ellipse is not selected, although its strokes may indeed be longer than the threshold. To ensure that such strokes are selected *and* connected to the main network, an approach that extracts shortest paths within the network is used. Specifically, if there is a stroke within a linear pattern that is longer than the threshold but not connected to a mesh that was retained in the first step, the shortest path in the source database (TLM3D) is computed between this stroke and the nearest *areal* segment selected in the first step of the integrated approach. This path is then selected for the target scale. The snippet on the right in Figure 6 shows the corresponding result. All the long strokes are selected and connected to the main network, whereas the dense mesh has been dissolved.

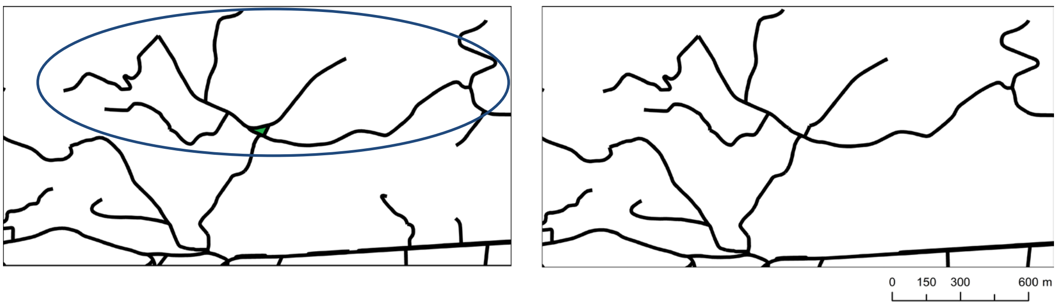

Figure 6. Linear patterns connected to a single dense mesh (left) and the result of the approach using the concept of shortest paths (right). Data: TLM3D © swisstopo.

4.3 Roundabout correction

Road networks usually contain roundabouts. These form dense meshes and the corresponding segments therefore are prone to elimination. As a consequence, insufficient results occur as depicted in Figure 7. One segment of the roundabout is eliminated (coloured in red), while the others are selected. However, a desirable result selects (or removes) the whole roundabout. Therefore, if only parts of a roundabout were selected using the integrated approach, all the other segments of the roundabout should be selected afterwards as well. To do that, however, the roundabouts first have to be identified and extracted.

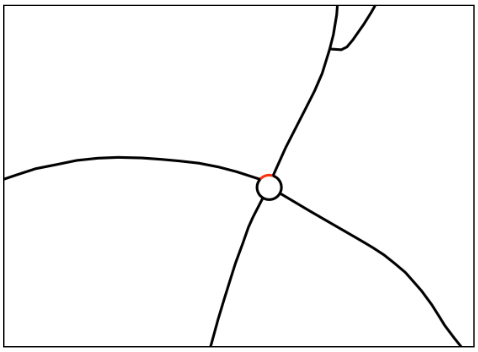

Figure 7. Insufficient result of a roundabout mesh elimination. Data: TLM3D © swisstopo.

To detect roundabouts, a stroke building algorithm can be used. Figure 8 shows a road network snippet with a typical roundabout, where the strokes are coloured differently. It can be seen that the roundabout results in a single stroke with a loop. This is due to the fact that the segments of the roundabout usually have a small deflection angle and therefore are concatenated into a single, circular stroke. Thus, in order to find all the roundabouts, one has to look for loops inside the strokes. A few additional parameter settings are necessary in order not to detect strokes that build loops but are no roundabouts, e.g. to only look for loops with a length smaller than a certain threshold (e.g. 100 m). This approach, developed together with Weiss and Weibel (2013), was able to detect all roundabouts in the four test areas.

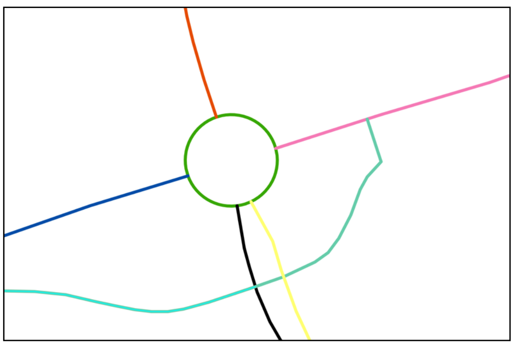

Figure 8. Roundabout and extracted strokes. Data: TLM3D © swisstopo.

4.4 Ensuring the accessibility of points of interest

Most approaches for an automated selection of road networks rely on the road network alone. However, the quality of the result might be improved if other feature classes were incorporated. One such feature class is a layer of points of interest (POIs). POIs are entities modelled as points having a certain importance. Examples are tramway stations, restaurants in touristic areas, sports arenas, or hospitals.

Depending on the target scale, POIs need to be accessible even in the generalised network. Figure 9 shows a snippet of an overlay of the source database (red) and the result of the integrated approach (black) without and with considering POIs (left and right snippet, respectively). The blue points represent shopping malls. As can be seen, they are not directly accessible in the snippet on the left. The following approach describes a mechanism to keep the POIs accessible in a generalised network.

After using the integrated approach, it can be checked whether the segment that is nearest to the POI in the source database (TLM3D) was selected. If so, nothing needs to be done, the POI is either accessible or it was not accessible in the source database in the first place. If the nearest segment is not selected, the POI might not be accessible. In this case, the shortest path in the source database is computed from the segment in the source database nearest to the POI and the segment in the *generalised* database nearest to the POI. This path can then be selected for the target scale. The snippet on the right in Figure 9 shows the result of this approach. The POIs now remain accessible. One could argue that the shortest path might not be the most important path regarding the usage. However, if the target scale is relatively large (in this case 1:50'000), the results showed that the selection of the shortest path is applicable, because it usually does not contain many segments. Additionally, there are not a large number of other paths that would connect the segment nearest to the POI to the main network. For smaller target scales, however, the shortest path might not be the best solution, since there could be other paths that are more important from a topological standpoint.

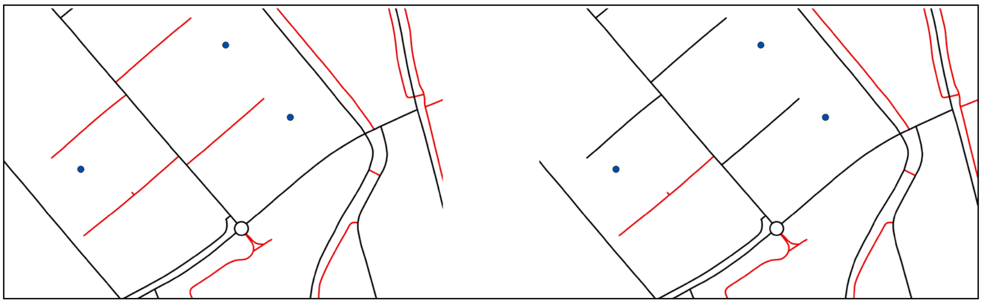

Figure 9. Inaccessible POIs (left) and accessible POIs (right). Source database segments omitted by the algorithm are coloured in red. Data: TLM3D © swisstopo.

4.5 Road network density in settlement areas

The first step in the integrated approach of Li and Zhou (2012) continuously thins out dense meshes by eliminating areal segments and merging adjacent meshes. Usually, dense meshes are located in urban settlement areas (Figure 10, left snippet). The settlement area in the northeast has a high number of dense meshes. The consequence is that the settlement areas are generalised to a greater extent than the surrounding rural areas and the density of the meshes is more or less evenly distributed in the result (Figure 10, right snippet). The settlement areas have been thinned out to such an extent that they are no longer recognisable as such. However, cartographic practice requires the main characteristics and structure of the original network to be retained.

Figure 10. Input database (left) and result without considering settlement areas (right). Data: TLM3D © swisstopo.

In the approach by Li and Zhou (2012) the global parameter based on which a mesh is eliminated and merged with an adjacent mesh is the mesh density. A natural line of thought therefore is to subtract a constant (denoted here as mesh density factor) from a mesh's density if it is in a settlement area. This ensures that an areal segment of a mesh in a settlement area is less likely to be eliminated. One way to decide whether a mesh is in a settlement area is to use an additional feature class that contains the settlement areas. The road network snippet on the left in Figure 11 shows the settlement area layer used for generating the result shown on the right. The settlement areas now remain clearly recognizable. The main structure is maintained. As a consequence, the surrounding rural areas are pruned to a greater extent. However, this cannot be avoided if the requirement is to eliminate a fixed amount (e.g. 30 %) of the segments from the source database. Thus, it is necessary to find a proper balance of segment elimination in urban and rural areas, respectively.

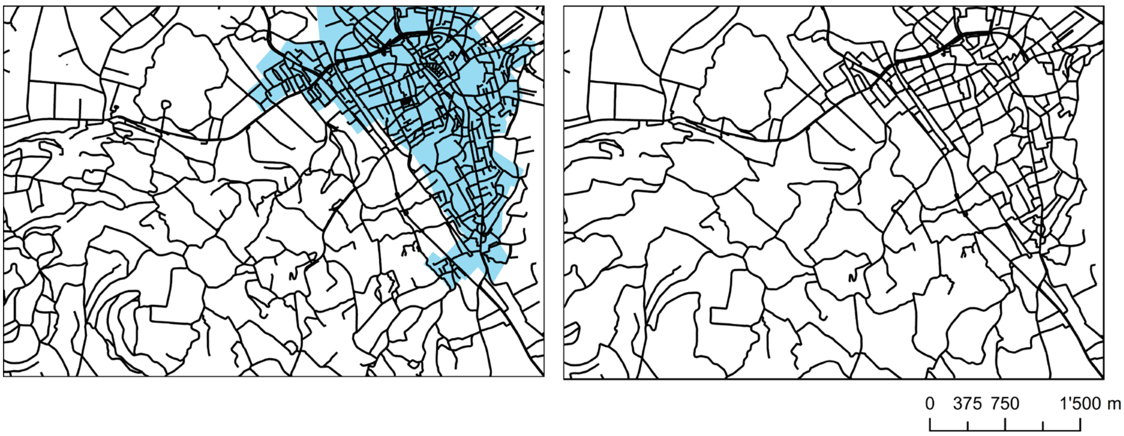

Figure 11. Settlement areas layer (left) used to generate a proper balance of segment elimination in urban and rural areas (right). Data: TLM3D © swisstopo.

A model that provides a quantitative measure to achieve a proper balance (i.e. to find an optimal mesh density factor) is based on the ratio of the number of segments in urban and rural areas. Figure 12 shows how this ratio depends on the mesh density factor. The red line represents the ratio of the input database, whereas the blue points show the ratio in the generalised result based on the chosen mesh density factor. The model now chooses the optimal mesh density factor retaining the ratio from the input database in the generalised result. Using this approach, the main structure can be maintained. As can be seen in Figure 12, the linear regression line fits the data almost perfectly ($R^2 = 0.99$).

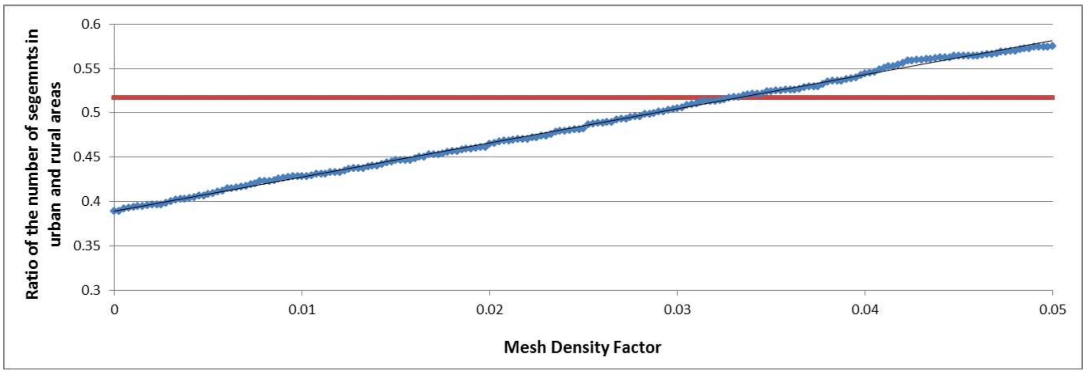

Figure 12. Ratio of the number of segments in urban and rural areas. The red line represents the ratio of the input database, whereas the blue points show the ratio in the generalised result based on the mesh density factor.

For the case where no settlement layer is available, a density algorithm was developed that extracts settlement areas (i.e. segments in settlement areas) from the road network itself. To decide whether a segment is in a settlement area or not, the approach calculates the centroid for each segment. If for a given centroid, the number of centroids within a certain radius exceeds a predefined threshold, the centroid (i.e. the segment) is considered to be in a settlement area. A few additional parameter settings are used, e.g. to omit segments in the calculation with certain attributes (e.g. highways) for the purpose of optimising the results. Figure 13 shows the result of that approach for one of the test areas. The settlement areas from the settlement layer are coloured in yellow. The segments identified to be part of settlement areas using the afore-mentioned detection algorithm are coloured in blue.

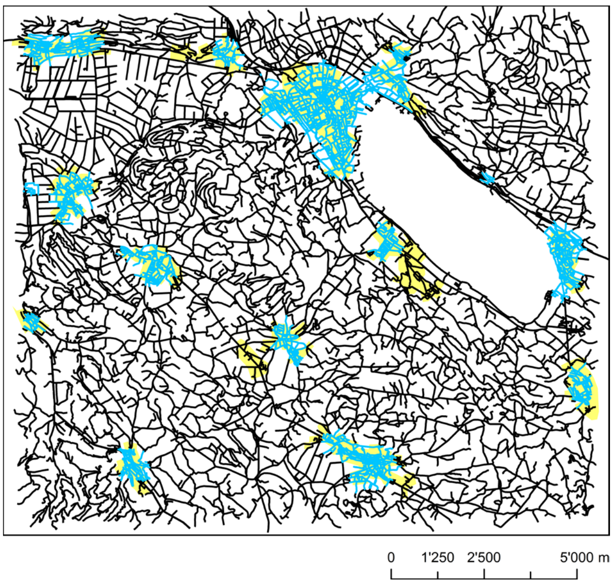

Figure 13. Result of the road segment density algorithm. The detected segments are coloured in blue, whereas the actual settlement boundaries are represented in yellow.

5. Conclusions and future work

This paper reported on a research project that started off by analysing three basic algorithms for road network selection (stroke-based, mesh-based and integrated approach) to find out which approach produces the most appropriate selection of the road network for a target scale of 1:50’000. While it could be shown that the integrated approach generates the most feasible results, several difficulties remained and diverse requirements set by swisstopo experts could not be met. Thus, the approach was extended by various concepts and additional feature classes were included. Quantitative evaluations, where the results were analysed with respect to the defined requirements, as well as a detailed evaluation by swisstopo experts revealed that these extensions improve the quality of the results significantly.

The approach was developed for a larger target scale of 1:50’000. Further research could reveal whether this approach is also applicable for smaller target scales. However, when generalising to smaller target scales, one also has to consider the larger structure and topological characteristics of the road network, e.g. using a centrality based approach as shown in Weiss and Weibel (2013). Furthermore, it could be analysed how the extended approach performs if the aim is not to select a certain percentage of segments for the target scale, but to select a total amount of road length.

Acknowledgements

We would like to thank swisstopo, in particular Dominik Käuferle, Stefan Wullschleger, Marianne Berger and Daniel Josi, for providing the data, advice and detailed feedback in the evaluation phase. Thanks are also due to Roy Weiss for conceptual discussions regarding the programming part.

References

Chaudhry O and Mackaness WA, 2005, Rural and Urban Road Network Generalisation Deriving 1:250,000 from OS MasterMap. In: *Proceedings of the 22th International Cartographic Conference*, La Coruña, Spain.

Chen J, Hu Y, Li Z, Zhao R and Meng L, 2009, Selective omission of road features based on mesh density for automatic map generalization. *International Journal of Geographical Information Science*, 23(8):1013-1032.

Jiang B and Claramunt C, 2004, A Structural Approach to the Model Generalization of an Urban Street Network. *GeoInformatica*, 8(2):157-171.

Li Z and Zhou Q, 2012, Integration of linear and areal hierarchies for continuous multi-scale representation of road networks. *International Journal of Geographical Information Science*, 26(5):855-880.

Liu X, Zhan FB and Ai T, 2010, Road selection based on Voronoi diagrams and “strokes” in map generalization. *International Journal of Applied Earth Observation and Geoinformation*, 12(2):194-202.

Mackaness WA, 1995, Analysis of Urban Road Networks to Support Cartographic Generalization. *Cartography and Geographic Information Systems*, 22(4):306-316.

Mackaness WA and Beard MK, 1993, Use of Graph Theory to Support Map Generalization. *Cartography and Geographic Information Systems*, 20(4):210-221.

Morisset B and Ruas A, 1997, Simulation and Agent Modelling for Road Selection in Generalisation. In: *Proceedings of the ICA 18th International Cartographic Conference*, Stockholm, Sweden, 1376-1380.

Thomson RC and Richardson DE, 1999, The ‘Good Continuation’ Principle of Perceptual Organization Applied to the Generalization of Road Networks. In: *Proceedings of the 19th International Cartographic Conference*, Ottawa, Canada, 1215-1223.

Touya G, 2010, A Road Network Selection Process Based on Data Enrichment and Structure Detection. *Transactions in GIS*, 14(5):595-614.

Weiss R and Weibel R, 2013, Road Network Selection for Small-Scale Maps Using an Improved Centrality Approach. In: *16th ICA Workshop on Generalisation and Map Production*, Dresden, Germany.

Yang B, Luan X and Li Q, 2011, Generating hierarchical strokes from urban street networks based on spatial recognition. *International Journal of Geographical Information Science*, 25(12):2025-2050.